

### IMAGE CLASSIFICATION

The present invention relates to an apparatus and method for classifying images and in particular for classifying elements within images. The invention is particularly, but not exclusively, useful for classifying pixels within hyperspectral images within the optical and non-optical domain.

5       The classification of spectral signatures in hyperspectral imagery is used for the identification of land cover types and may be used for the identification of specific target objects of interest where their spectral characteristics are known. The typical approach to this type of classification problem uses a set of "training data" to characterise the statistical distributions of regions ("classes") of known 10 land cover type. These class distributions may then, in turn, be used to recognise previously unseen samples of the same type of data, the latter samples being assigned to one of the classes of training data.

15      The major problem with this approach is that a large number of training samples of each class type are typically needed to completely characterise the statistical distribution of each of the classes. Thus a very large training dataset needs to be assembled. The assembly of a training dataset for hyperspectral imagery is usually done by carrying out data collection trials in the field; an expensive and time-consuming operation.

20      In recent times a number of new statistical techniques have been developed which reduce the volume of training data required, at the expense of considerably increased complexity in the classification process. An example of such a technique is described by Skurichina, M and Duin, R. P. W., in "Bagging and the Random Subspace Method for Redundant Feature Spaces", 25 Proceedings of the 2<sup>nd</sup> International Workshop on Multiple Classifier Systems, Cambridge, UK, pp 1-10, July 2001. One such technique, the Random Subspace Method (RSM), has been successfully applied, as described by Willis, C. J., in "Classification of Hyperspectral Imagery using Limited Training Data Samples", Proceedings of SPIE, Image and Signal Processing for Remote 30 Sensing VIII, 4885, pp 379-388, 2003, to hyperspectral data allowing a considerable reduction in the volume of training data required for only a modest

- 2 -

reduction in classification performance. The RSM builds an ensemble of classifiers each based on a different view of the training dataset. The output of each member of the classifier ensemble, when applied to new sample data, is combined to produce the ensemble classification. It is normal for the  
5 combination method to be a majority vote method.

The approach taken by the RSM is to select, at random, a subset of the features of the full problem and to use these features alone to train one of the "basis" classifiers used in the ensemble. If a large number of basis classifiers are trained in this way, then it is possible that the ensemble will have a superior  
10 performance to that of a single classifier trained on the full feature space. This has been found to be the case in a number of application domains.

An additional benefit of this approach relates to its use on small training datasets. If the size of the training dataset is smaller than the dimensionality of the original problem, then the class statistics become either difficult or  
15 impossible to estimate and it may turn out to be impossible to use the chosen decision rule of the basis classifiers. By restricting the size of the feature space for each basis classifier, such that the class statistics for each ensemble element are calculable, then it becomes possible to produce classifications in this difficult case.

In the RSM, the set of features are selected randomly for each ensemble  
20 basis classifier. To ensure that at least most of the available features are used, a large number of basis classifiers must be used in the ensemble. Referring to Figure 1, an example is shown of a simple classifier designed according to the RSM and having an ensemble of only four basis classifiers. However, the use of  
25 a large number of basis classifiers results in a significant computational requirement when using the method which, in turn, can make the RSM unattractive to use in time-critical applications.

Another example of a known subspace selection method is the "Classical Feature Extraction" method, described for example by Fukunaga, K., in the  
30 book "Introduction to Statistical Pattern Recognition", Second Edition, Academic Press, 1990. In this method, much of the processing is carried out

- 3 -

offline to select the combination of features from the feature space most likely to ensure class separability. Only the selected subset of each feature vector for elements to be classified is then input to a single classifier with relatively low operational processing requirements. However, the selection technique in the  
5 classical feature selection method is, to a large extent, based on the statistical properties of the available training data and may therefore suffer from the same problems as the classifiers themselves when training datasets are small. That is, the poor estimation of class mean vectors, covariance matrices or scatter matrices can, in turn, lead to poor estimates of the set of discriminatory  
10 features.

As sensor technology develops, the quantity of data that can be made available to image classification systems is ever increasing. Techniques with a large processing requirement are therefore likely to be of limited application for some time to come if the full range of available sensor data is to be exploited.

15 From a first aspect, the present invention resides in an apparatus for classifying elements, in particular elements within an image, wherein an element is defined by a vector of feature values, the apparatus comprising:

classifier means comprising a plurality of classifiers each operable, in respect of an element to be classified, to receive a different predetermined  
20 subset of the feature values from the element feature vector and wherein, in operation, each said classifier is trained in respect of a predetermined set of classes using training data representative of elements in each said class; and

25 combining means operable to combine outputs from the plurality of classifiers to determine which of the predetermined classes to associate with an element to be classified,

characterised in that each of said different predetermined subsets of feature values comprise a different cyclic selection of the feature values such that, in operation, adjacent feature values in an element feature vector are input to different ones of said plurality of classifiers and all feature values are input to  
30 at least one classifier.

- 4 -

Features may be selected cyclically according to "round robin" basis. As such, the subspace selection technique embodied in preferred embodiments of the present invention will be referred to as the "structured subspace method".

Preferred embodiments of the present invention therefore approach the problem of distributing closely matched features in the feature space across an ensemble of basis classifiers in a structured manner, so greatly reducing the number of classifiers required while still making use of the full feature space available.

In some applications it may be appropriate for preferred embodiments of the present invention to be used to provide initial indications of a class of object in a image and for a further classifier, designed according to the random subspace approach for example, to be used to further refine the classification of that object where time is not so critical.

Majority voting is the preferred technique by which the output of basis classifiers may be combined to produce a classification decision, although other forms of voting, such as posterior probability, may be used.

By way of an example of the type of image to which preferred embodiments of the present invention may be applied is the well-known AVIRIS Indian Pines image (Landgrebe, D. A., Biehl, L., "AVIRIS Indian Pines Reflectance Data: 92AV3C", available as a part of the documentation for the MultiSpec hyperspectral imagery analysis environment at the internet address <http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html>), a largely agricultural scene containing some difficult to separate classes of ground cover.

From a second aspect, the present invention resides in a method for classifying elements, in particular elements within an image, wherein an element is defined by a vector of feature values, the method comprising the steps of:

(i) using, for each a set of predetermined classes, a training dataset representative of elements in the class to train a plurality of classifiers in respect of the class, wherein each classifier is operable to receive feature vector values in respect of a different predetermined cyclic selection of features such that

- 5 -

adjacent feature values in an element feature vector are input to different ones of said plurality of classifiers and all feature values are input to at least one classifier;

(ii) receiving a feature vector for an element to be classified;

5 (iii) inputting the received feature vector values to said plurality of trained classifiers according to said predetermined cyclic selections and generating a plurality of classifier outputs; and

(iv) combining the classifier outputs to determine which of said predetermined classes to associate with the element to be classified.

10 A preferred embodiment of the present invention will now be described by way of example only and with reference to the accompanying drawings, of which:

Figure 1 shows an example of a known classifier based upon random subspace selection as discussed above; and

15 Figure 2 shows an example of a classifier according to a preferred embodiment of the present invention.

A known classifier designed according to the known random subspace method (RSM), as discussed above, will firstly be summarised with reference to Figure 1.

20 Referring to Figure 1, a feature vector 100 is shown comprising all features of the available feature space. In the example of a hyperspectral image classifier, the feature vector 100 representing an element of an image to be classified comprises a vector of intensity values for each of the frequency bands of the image.

25 According to the RSM, the features represented by the feature vector 100 are associated in a random manner with an ensemble of basis classifiers 105 such that the number of features input to each basis classifier 105 – the subspace dimension - is the same. However, as can be seen from Figure 1, because the selection of features for each basis classifier 105 is random, not all

- 6 -

features are necessarily selected for consideration by the ensemble in classifying a given element feature vector 100. The best that can be achieved is to provide a sufficient quantity of basis classifiers 105 in the ensemble so that the probability of selection of any one feature is at least a predetermined figure, e.g. 99%. Clearly, the higher the figure, the greater the number of basis classifiers 105 that need to be provided in the ensemble.

The results from each basis classifier 105 of the ensemble in respect of an element to be classified are input to a vote 110 where the results are combined by a majority vote to determine the classification result.

A particular disadvantage of the classifier of Figure 1 is the high level of processing required to train and operate the classifiers 105 given their large number.

A preferred embodiment of the present invention will now be described with reference to Figure 2. Features of Figure 2 in common with Figure 1 are labelled with the same reference numerals.

Referring to Figure 2, a feature vector 100 defining an element to be classified is shown, spanning the available feature space as for the classifier of Figure 1. However, in the preferred method of the present invention, a structured approach is taken to the association of features of the feature vector 100 with each of a predetermined number of basis classifiers 105, in this example with two basis classifiers 105. This approach guarantees that all the features of a feature vector 100 are considered by the ensemble of basis classifiers 105 while ensuring also that, where adjacent features are closely related, they are distributed amongst the classifiers 105 in the ensemble.

Preferably, features are associated with each of the basis classifiers 105 using a cyclic, or "round-robin" selection. In the specific example of Figure 2 having two basis classifiers 105, features are associated alternately with one classifier 105 then the other throughout the length of the feature vector 100 until all features are assigned. As for Figure 1, the results of the trained classifiers 105 are combined in a vote 110 to determine the classification results for a given element feature vector 100.

- 7 -

Where the number of classifiers does not exactly divide the number of features, elements of the feature vector 100 may be reused such that all basis classifiers 105 have the same dimensionality. This approach guarantees that all elements of the feature vector are assigned to at least one basis classifier 105 and, for a given subspace dimensionality, a significantly smaller number of basis classifiers 105 is required to span the available feature space, in comparison with a classifier designed according to the RSM, with consequent savings on processor loading during training and operation.

Although a preferred embodiment of the present invention has been discussed in the context of hyperspectral image classification, it will be clear that a feature vector 100 defining an element of an image to be classified need not relate to bands of optical frequencies as in hyperspectral images, but may relate to other types of feature in an "image" by which elements may be defined and classified. The word "image" is used broadly in the present patent specification to mean not only an optical image where, for example, features may represent the intensity of a pixel in each of a number of optical frequency bands, but also an image defined in terms of other feature parameters, for example those characterising an image generated using magnetic resonance interferometry (MRI) or other "imaging" technique.

As mentioned in the introductory part of the present patent specification, the preferred embodiment of the present invention is an example of a selected subspace method in which an ensemble of classifiers is assembled. The underlying, or "basis" classifiers used in the ensemble may be of any one of a number of known types. For example, the basis classifiers may be of a type known as a quadratic Bayes classifier, described for example in the book by Fukunaga, referenced above, with slight modifications required to deal with singular covariance matrices, i.e. if a class conditional covariance matrix is found to be ill-conditioned it is replaced by the common covariance matrix of all classes; if the common covariance matrix is also found to be ill-conditioned then its diagonal only is used.

- 8 -

An alternative choice of basis classifier is a neural network. The choice of classifier is not therefore an essential feature of the present invention and will not be described further in this patent specification.

In practice, for example using the data used is a part of a scene collected by the Airborne Visual and near infra red (IR) Imaging Spectrometer (AVIRIS) referenced above, it has been found that there may be considerable correlation between neighbouring elements of the feature vector 100. The structured subspace method of the present invention advantageously disperses these correlated elements throughout the classifier ensemble, thereby ensuring that each basis classifier 105 is, individually, a good subspace classifier. An ensemble built from such a collection might be expected to improve on the performance in respect of any individual element to be classified.

In practice, it has been found that the structured subspace approach of the present invention method closely follows the performance of the random subspace method. However, while the latter may often be able to deliver a marginally better peak performance, it is at a considerably higher computational cost.

Both the known random subspace ensemble method and the structured subspace method of the present invention have been found applicable to difficult classification problems for pixels in hyperspectral imagery. The present set sizes are small compared to the dimensionality of the problem. The present structured subspace method is able to produce results very close to those achievable using the random subspace method, but using a significantly smaller ensemble of basis classifiers 105, and therefore at a significantly reduced computational cost.

The invention has been described, by way of example only, and it will be appreciated that variation may be made to the embodiment described without departing from the scope of invention. For example the invention may be employed in spectroscopy, in classifying pixels within images obtained from imaging equipment such as digital cameras, charge coupled devices (CCDs),

- 9 -

magnetic resonance imagers (MRI) or other imaging devices operating at optical and other wavelengths. The present invention may also be used in novelty identification and in a range of applications in which a large amount of sensor data, across a broad waveband, needs to be assessed for classification  
5 quickly and efficiently.